

George G. Vega Yon, Ph.D.

mo: (626) 381-8171 | email: g.vegayon@gmail.com | Sandy, UT, 84094

<https://ggv.cl> / <https://github.com/gvegayon> / <https://linkedin.com/in/georgevegayon> / <https://ggv.cl/scholar>

PROFESSIONAL SUMMARY

Senior Data Scientist & Research Engineer with 15+ years of experience in **high-performance computing (HPC)**, **large-scale simulations**, and **scientific software development**. Expert in designing and optimizing **R/Python packages**, **scalable data pipelines**, and **cloud/HPC workflows**, with a proven track record of leading technical teams to deliver **production-ready solutions**. Creator of open-source tools with **1M+ downloads**, widely adopted in academia, government, and industry.

EXPERIENCE

Assistant Professor of Data Science, The University of Utah

Nov 2021 – Present

- Designed advanced frameworks in R/C++ for large-scale epidemic and network modeling on HPC clusters.
- Created and maintained widely adopted open-source statistical software (1M+ downloads) for epidemiology, optimization, and network analysis.
- Directed interdisciplinary teams to produce production-ready code, peer-reviewed publications, and decision-support tools for federal agencies.
- Designed the first graduate-level course on HPC with R/C++, advancing computational data science training.
- Core faculty at the Utah Center for Data Science + AI, driving campus-wide initiatives in advancing DS+AI.

Data Scientist (Contractor), Centers for Disease Control and Prevention

May 2024 – Present

- Supported scientific software development by building and maintaining R packages and Python libraries for epidemiological research.
- Designed and deployed software engineering infrastructure (CI/CD pipelines, containerized workflows, reproducible environments).
- Applied network modeling and agent-based simulations to infectious disease transmission, outbreak preparedness, and policy support.

Lead Data Scientist (Associate), Booz Allen Hamilton

Nov 2023 – May 2024

- Led development of a new epidemiological forecasting framework using Python and probabilistic programming.
- Directed market research on HPC and cloud technologies for large-scale scientific applications.
- Analyzed government open-source software ecosystems by mining millions of GitHub code lines to evaluate adoption and innovation trends.

Research Programmer II, University of Southern California

Feb 2018 – Nov 2021

- Engineered R packages and HPC workflows for machine learning and network science research, improving runtime performance and reproducibility.
- Taught graduate-level course "Intro to Health Data Science," bridging research methods with modern data science tools.
- Contributed to \$10M+ in funded projects by providing software engineering expertise, statistical modeling, and scalable data pipelines.

Programmer Analyst II, University of Southern California

Oct 2015 – Feb 2018

- Served as a consultant in statistics and computer science across campus.
- Founder of the "R Bootcamp for Statistical Computing."
- Wrote scientific papers and software on network science and presented them at conferences.
- Designed and led workshops on R and Social Network Analysis.

EDUCATION

Ph.D. in Biostatistics, University of Southern California, 2020

M.Sc. in Economics, California Institute of Technology, 2016

MA in Economics and Public Policy, Universidad Adolfo Ibáñez, 2011

SKILLS

R, C++, SQL, Python, XML, NLP, Stata, AWS, Git, GitHub, Docker/Podman, TensorFlow, PyTorch, Deep Learning, CI, Slurm, Unix, Jira, Scrum, Kanban, team management and coordination, R Shiny, excellent communication skills.

SOFTWARE PACKAGES (selected)

- *epiworldR: Fast Agent-Based Epi Models* (2023). R package version 0.2-1 URL: <https://cran.r-project.org/package=epiworldR> .
- *rgexf: Build, Import and Export GEXF Graph Files* (2020). R package version 0.16.0. URL: <https://CRAN.R-project.org/package=rgexf>.
- *netdiffuseR: Analysis of Diffusion and Contagion Processes on Networks* (2020). R package version 1.22.0. URL: <https://cran.r-project.org/package=netdiffuseR>.
- *ergmito: Exponential Random Graph Models for Small Networks* (2020). R package version 0.3-0. URL: <https://cran.r-project.org/package=ergmito>.
- *slurmR: A Lightweight Wrapper for 'Slurm'* (2020). R package version 0.4-1. URL: <https://CRAN.R-project.org/package=slurmR>.
- *fmcmc: A friendly MCMC framework* (2020). R package version 0.3-0. URL: <https://CRAN.R-project.org/package=fmcmc> .

ACADEMIC PUBLICATIONS (selected)

- Sima Najafzadehkhoei, George G. Vega Yon, Bernardo Modenesi, and Derek S. Meyer. “*Machine Generalize Learning in Agent-Based Models: Going Beyond Surrogate Models for Calibration in ABMs.*” In: **arXiv preprint** arXiv:2509.07013 (2025). URL: <https://arxiv.org/abs/2509.07013>
- George G. Vega Yon. “*Power and Multicollinearity in Small Networks: A Discussion of “Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks” by Krivitsky, Coletti & Hens.*” In: **Journal of The American Statistical Association** (2023). to appear.
- George G. Vega Yon, Andrew Slaughter, and Kayla de la Haye. “*Exponential random graph models for little networks.*” In: **Social Networks** 64 (2021), pp. 225–238. URL: <https://doi.org/10.1016/j.socnet.2020.07.005>.
- George G. Vega Yon, Duncan C. Thomas, John Morrison, Huaiyu Mi, et al. “*Bayesian parameter estimation for automatic annotation of gene functions using observational data and phylogenetic trees.*” In: **PLOS Computational Biology** 17.2 (Feb. 2021), pp. 1–35. URL: <https://doi.org/10.1371/journal.pcbi.1007948>.
- George G. Vega Yon and Paul Marjoram. “*fmcmc: A friendly MCMC framework.*” In: **Journal of Open Source Software** 4.39 (July 2019), p. 1427. URL: <http://joss.theoj.org/papers/10.21105/joss.01427>.
- George G. Vega Yon and Brian Quistorff. “*parallel: A command for parallel computing.*” In: **The Stata Journal: Promoting communications on statistics and Stata** 19.3 (Sept. 2019), pp. 667–684. URL: <http://journals.sagepub.com/doi/10.1177/1536867X19874242>.

AWARDS

- Best Paper Awards (Computational Methods) 72 ICA conference, 2022.
- Travel Grant, Society of Young Network Scientist, 2019.
- Fellowship, California Institute of Technology, 2014.
- Scholarship, Adolfo Ibáñez University, 2006.

REFEREE (ad hoc reviewer)

Journal of The American Statistical Association, American Sociological Review, BMC Infectious Diseases, The Official Journal of The Society for Computational Economics, The R Journal, Social Networks, Journal of Mathematical Sociology, Journal of Open Source Software, Bioinformatics, Computer Methods and Programs in Biomedicine Update